



Nonparametric FBST for Validating Linear Models [†]

Rodrigo F. L. Lassance ^{1,2,*} , Julio M. Stern ³ and Rafael B. Stern ³ ¹ Department of Statistics, Federal University of São Carlos, São Paulo 13565-905, Brazil² Institute of Mathematics and Computer Sciences, University of São Paulo, São Paulo 13566-590, Brazil³ Institute of Mathematics and Statistics, University of São Paulo, São Paulo 05508-090, Brazil; jstern@ime.usp.br (J.M.S.); rbstern@gmail.com (R.B.S.)

* Correspondence: rflassance@gmail.com

[†] Presented at the 43rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Ghent, Belgium, 1–5 July 2024.

Abstract

In Bayesian analysis, testing for linearity requires placing a prior to the entire space of potential regression functions. This poses a problem for many standard tests, as assigning positive prior probability to such a hypothesis is challenging. The Full Bayesian Significance Test (FBST) sidesteps this issue, standing out for also being logically coherent and offering a measure of evidence against H_0 , although its application to nonparametric settings is still limited. In this work, we use Gaussian process priors to derive FBST procedures that evaluate general linearity assumptions, such as testing the adherence of data and performing variable selection to linear models. We also make use of pragmatic hypotheses to verify if the data might be compatible with a linear model when factors such as measurement errors or utility judgments are accounted for. This contribution extends the theory of the FBST, allowing for its application in nonparametric settings and requiring, at most, simple optimization procedures to reach the desired conclusion.

Keywords: FBST; HPD; Bayesian nonparametrics; linear model; Gaussian process; pragmatic hypothesis

1. Introduction

Although linear models are widespread in the scientific literature, their validity is rarely tested in its full complexity. Generally, linearity is tested as a particular case of a more general parametric model [1] or compared to a finite selection of models—each with their own prior specification—through measures such as the Deviance Information Criterion [2]. In actuality, testing the adherence of linear models to data requires (i) assigning a nonparametric prior to the set of regression functions and (ii) devising a procedure that highlights the evidence against the linear model hypothesis based on the data and the prior. Devising a test solely based on the posterior probability of the hypothesis in this case is seldom advised, as it imposes positive prior probability to the set of linear models when there are countless nonlinear functions arbitrarily close to any element of it.

The Full Bayesian Significance Test (FBST, [3]) is the testing framework used throughout this work. The FBST does not violate the likelihood principle, does not require setting positive prior probabilities to hypotheses, and provides a measure of evidence against H_0 , along with other desirable characteristics. With the exceptions of Corrêa Filho [4] and Liu et al. [5], the FBST has not been applied to nonparametrics, still requiring new theoretical developments to systematically embrace such settings.



Academic Editor: Geert Verdoolaege

Published: 24 September 2025

Citation: Lassance, R.F.L.; Stern, J.M.; Stern, R.B. Nonparametric FBST for Validating Linear Models. *Phys. Sci. Forum* **2025**, *12*, 2. <https://doi.org/10.3390/psf2025012002>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Bridging the gaps above, this paper provides a nonparametric FBST formulation that tests the adherence of linear models to data. By using a Gaussian Process (GP, [6]) as a prior to model the regression function, we propose FBST procedures that depend on whether the covariates' domain $\mathcal{X}$ is finite or infinite. Furthermore, we lay out FBST procedures for hypotheses that include negligible deviations from H_0 , known as pragmatic hypotheses [7], useful to evaluate if H_0 is approximately instead of precisely compatible with the data.

In Figure 1, we illustrate how the FBST operates when applied to H_0 and its pragmatic version, $Pg(H_0)$. For $\alpha \in (0, 1)$, the posterior is used to obtain the $(1 - \alpha)100\%$ Highest Posterior Density (HPD) region, the smallest credible region with probability $(1 - \alpha)$ of containing the quantity of interest. The hypothesis is rejected if it does not intersect with the HPD. This procedure is meaningful even when H_0 is precise, that is, when $\mathbb{P}(H_0) = 0$.

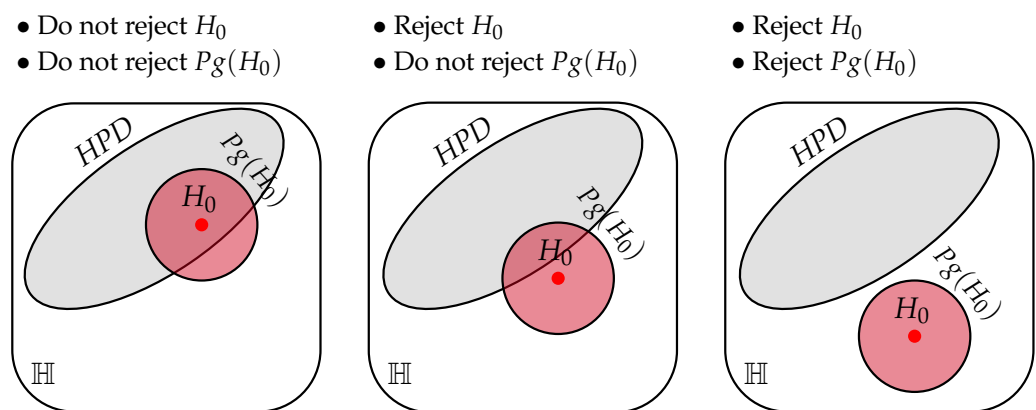


Figure 1. Illustration of the FBST for a precise H_0 and its pragmatic version, $Pg(H_0)$, in the hypothesis space $\mathbb{H}$. Each panel presents a possible configuration of the hypotheses and the HPD, with the text above the panels indicating the conclusion.

Even though our contribution makes exclusive use of the FBST, this does not imply that it is the only valid framework for the problem. While no other testing procedure as general as ours has been proposed, Mulder [8] uses Bayes factors to test if a single covariate may be nonlinearly related to the response variable and Lassance et al. [9] (Section 3.1) test the pragmatic version of the linear model hypothesis through its posterior probability.

This work is organized as follows. In Section 2, the required background knowledge is provided. Our findings are presented in Section 3, leading in Section 4 to an application that puts all the FBST procedures to use. Lastly, Section 5 describes how to enhance the FBST further and establishes potential future research. All proofs can be found in Appendix A.

2. Materials and Methods

2.1. Full Bayesian Significance Test (FBST)

The FBST is composed of three steps [3]. For $H_0 : \theta \in \Theta_0 \subset \Theta$, where Θ is the parameter space, these steps are as follows:

1. Delimit the set of elements in Θ that are more likely than those in Θ_0 . That is, if $f(\theta|\mathcal{D})$ is the posterior density of θ given the data $\mathcal{D}$,

$$T := \left\{ \theta \in \Theta : f(\theta|\mathcal{D}) \geq \sup_{\theta \in \Theta_0} f(\theta|\mathcal{D}) \right\}.$$

2. Obtain e-value $:= 1 - \mathbb{P}(\theta \in T|\mathcal{D}) = 1 - \int_T f(\theta|\mathcal{D})d\theta$, the Bayesian evidence value.
3. Reject H_0 if e-value $\leq \alpha$ for a previously specified significance level $\alpha \in (0, 1)$.

In this paper, we use a procedure equivalent to the FBST: reject H_0 if the Highest Posterior Density (HPD) region is such that $HPD \cap \Theta_0 = \emptyset$. The HPD is the smallest region with posterior probability of $1 - \alpha$, obtained by finding the value f^* such that

$$\mathbb{P}(\theta \in HPD|\mathcal{D}) = 1 - \alpha, \quad HPD := \{\theta \in \Theta : f(\theta|\mathcal{D}) \geq f^*\}.$$

When $\theta|\mathcal{D}$ is normally distributed, the HPD region is equivalent to the credible interval symmetric around the posterior mean. For its multivariate counterpart, $\theta|\mathcal{D} \sim N_k(\boldsymbol{\mu}, \Sigma)$, we have that $\Sigma^{-1/2}(\theta - \boldsymbol{\mu}) \sim N_k(\mathbf{0}, \mathbb{I})$, and thus $(\theta - \boldsymbol{\mu})'\Sigma^{-1}(\theta - \boldsymbol{\mu}) \sim \chi_k^2$, where χ_k^2 stands for the chi-squared distribution with k degrees of freedom. Therefore, if $q_{(1-\alpha)}(\cdot)$ is the $(1 - \alpha)100\%$ quantile function, the HPD is given by the following ellipsoid ([10], Result 4.7):

$$\{\boldsymbol{\theta} \in \mathbb{R}^k : (\boldsymbol{\theta} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}) \leq q_{(1-\alpha)}(\chi_k^2)\}. \quad (1)$$

2.2. Gaussian Processes (GPs)

A GP is a nonparametric family of priors used to model functions in regression settings. The random function $g : \mathcal{X} \rightarrow \mathbb{R}$ behaves according to a GP if

$$g(\mathbf{X}) \sim N(m(\mathbf{X}), K(\mathbf{X}, \mathbf{X})), \quad \forall \mathbf{X} \subset \mathcal{X},$$

where $m(\cdot)$ and $K(\cdot, \cdot)$, respectively, determine the mean and covariance of the process. When the response variable Y is such that $Y = g(\mathbf{x}) + \epsilon$ for $\epsilon \sim N(0, \sigma^2)$, that is,

$$L(g, \sigma^2|\mathbf{y}, \mathbf{X}) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - g(\mathbf{x}_i))^2\right\},$$

then the GP is conjugate and its posterior is such that

$$\begin{aligned} g(\mathbf{X}')|\mathbf{y}, \mathbf{X}, \sigma^2 &\sim N(\boldsymbol{\mu}(\mathbf{X}'), \Sigma(\mathbf{X}', \mathbf{X}')), \\ \boldsymbol{\mu}(\mathbf{X}') &:= m(\mathbf{X}') + K(\mathbf{X}, \mathbf{X}')(K(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbb{I})^{-1}(\mathbf{y} - m(\mathbf{X})), \\ \Sigma(\mathbf{X}', \mathbf{X}') &:= K(\mathbf{X}', \mathbf{X}') - K(\mathbf{X}, \mathbf{X}')(K(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbb{I})^{-1}K(\mathbf{X}', \mathbf{X}). \end{aligned}$$

The choice of m , K , and σ^2 reflect positions on the mean, smoothness, and variation surrounding the GP. In Section 4, we use the specifics of the application to choose them. For more general settings, one may assume that the uncertainty of m and K is reducible to a finite number of parameters. Then, one can either set priors to such parameters directly [11] or plug point estimates for them based on the maximum partial likelihood [12].

Conditionally on σ^2 , the HPD region of the GP can be analytically obtained for any finite set $\mathbf{X}' = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)'$. Since the marginals of the posterior GP are also normally distributed, Equation (1) entails that the $(1 - \alpha)100\%$ HPD region for $g(\mathbf{X}')|\mathbf{y}, \mathbf{X}$ is

$$\{h \in \mathbb{H} : (h(\mathbf{X}') - \boldsymbol{\mu}(\mathbf{X}'))'\Sigma(\mathbf{X}', \mathbf{X}')^{-1}(h(\mathbf{X}') - \boldsymbol{\mu}(\mathbf{X}')) \leq q_{(1-\alpha)}(\chi_m^2)\}. \quad (2)$$

It is also possible to obtain an HPD set for the GP without setting $\mathbf{X}'$. Let $\mathbb{P}_g$ and $\mathbb{P}_{g|\mathbf{y}, \mathbf{X}, \sigma^2}$, respectively, be the prior and posterior probability measures of the GP defined on a measurable space $(\mathcal{G}, \mathbb{G})$. Hence,

$$\mathbb{P}_{g|\mathbf{y}, \mathbf{X}, \sigma^2}(A) = \frac{\int_A L(g, \sigma^2|\mathbf{y}, \mathbf{X})d\mathbb{P}_g(h)}{\int_{\mathcal{G}} L(g, \sigma^2|\mathbf{y}, \mathbf{X})d\mathbb{P}_g(h)}, \quad \forall A \subset \mathbb{G}.$$

Since $\mathbb{P}_{g|y,X,\sigma^2} \ll \mathbb{P}_g$, the Radon–Nikodym derivative of the GP for $h \in \mathcal{G}$ is such that

$$\frac{d\mathbb{P}_{g|y,X,\sigma^2}}{d\mathbb{P}_g}(h) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - h(x_i))^2\right) = \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - h(\mathbf{X}))'(\mathbf{y} - h(\mathbf{X}))\right), \quad (3)$$

i.e., it suffices to evaluate h only on the values of $\mathbf{X}$ in the sample. To account for repeated lines in $\mathbf{X}$, let $\mathbf{X}^*$ be the matrix with all unique observations from $\mathbf{X}$ and n^* be the number of lines of $\mathbf{X}^*$. Defining $D_{n_{\mathbf{X}^*}}$ as a diagonal matrix that counts how many times each $\mathbf{x} \in \mathbf{X}^*$ appears in $\mathbf{X}$ and $\bar{\mathbf{y}}_{\mathbf{X}^*}$ as the vector of means of all elements of $\mathbf{y}$ related to each $\mathbf{x}$,

$$\frac{d\mathbb{P}_{g|y,X,\sigma^2}}{d\mathbb{P}_g}(h) \propto \exp\left(-\frac{1}{2\sigma^2} (h(\mathbf{X}^*) - \bar{\mathbf{y}}_{\mathbf{X}^*})' D_{n_{\mathbf{X}^*}} (h(\mathbf{X}^*) - \bar{\mathbf{y}}_{\mathbf{X}^*})\right). \quad (4)$$

Thus, for a constant c_α , a Weighted Residual Sum of Squares (WRSS) defines the HPD:

$$\text{HPD}_{(1-\alpha)} = \{h \in \mathcal{G} : \text{WRSS}(h) \leq c_\alpha\}, \text{WRSS}(h) := (h(\mathbf{X}^*) - \bar{\mathbf{y}}_{\mathbf{X}^*})' D_{n_{\mathbf{X}^*}} (h(\mathbf{X}^*) - \bar{\mathbf{y}}_{\mathbf{X}^*}). \quad (5)$$

2.3. Pragmatic Hypotheses

The pragmatic hypothesis enlarges H_0 to a set deemed as practically equivalent. The implementation uses the notion of negligible deviations from H_0 . The degree to which the hypothesis is enlarged depends on the choice of a threshold ε , and factors such as the scale of measurement errors or expert's utility judgments could help set it (see Section 4 for a practical example and (Lassance et al. [9], Section 4) for suggestions). Formally, for a hypothesis space $\mathbb{H}$, let $d(\cdot, \cdot)$ be the dissimilarity function from which one can express how much of a departure from H_0 is reasonable. Then, the pragmatic hypothesis is given by

$$Pg(H_0, d, \varepsilon) := \bigcup_{h_0 \in H_0} \{h \in \mathbb{H} : d(h_0, h) \leq \varepsilon\} = \left\{h \in \mathbb{H} : \inf_{h_0 \in H_0} d(h_0, h) \leq \varepsilon\right\}, \quad (6)$$

that is, the pragmatic hypothesis contains all elements $h \in \mathbb{H}$ such that, for some element $h_0 \in H_0$, $d(h_0, h) \leq \varepsilon$. In this work, we assume that $\mathbb{H} = \mathcal{G}$ is a space of functions of the type $h : \mathcal{X} \rightarrow \mathbb{R}$. Further specifications on $\mathbb{H}$ are presented in Section 3. When $d(\cdot, \cdot)$ and ε are implicit, we use $Pg(H_0)$ to denote the pragmatic hypothesis.

3. Results

Throughout this work, we use the modeling assumptions in Section 2.2 for the data $(\mathbf{y}, \mathbf{X})$ and the regression function $g(\cdot)$, and assume that the hypothesis of interest is

$$H_0 : g(\mathbf{x}) = \mathbf{b}(\mathbf{x})\boldsymbol{\beta}, \quad \forall \mathbf{x} \in \mathcal{X}, \quad \boldsymbol{\beta} \in \mathbb{R}^k, \quad (7)$$

where $\mathbf{b}(\mathbf{x}) = (b_1(\mathbf{x}), b_2(\mathbf{x}), \dots, b_k(\mathbf{x})) \subset \mathbb{H}$ is a linearly independent set of linear functions and $\mathcal{X}$ is the covariates' domain. The choice of $\mathbf{b}$ determines the test performed, such as evaluating linear models ($\mathbf{b}(\mathbf{x}) := \mathbf{x}$) or doing variable selection ($\mathbf{b}(\mathbf{x}) := \mathbf{x}_{-i}$).

Our findings are divided in two settings: those applicable to H_0 and those to $Pg(H_0)$. In both cases, we explore when $\mathcal{X}$ is a finite or an infinite set. The finite case provides a closed-form solution for the FBST of H_0 and a solution for the pragmatic hypothesis that requires a univariate optimization procedure. When $\mathcal{X}$ is infinite, testing H_0 or $Pg(H_0)$ also requires determining c_α in the HPD of Equation (5), which is achieved by noting that the WRSS can be expressed as a linear combination of noncentral chi-squared random variables; therefore, c_α is the $(1 - \alpha)$ quantile of a generalized chi-squared distribution [13].

Theorem 1 (FBST of the linear model hypothesis). Let H_0 be the hypothesis in Equation (7) and $g(\cdot)|\mathbf{y}, \mathbf{X} \sim GP(\mu(\cdot), \Sigma(\cdot, \cdot))$. Then,

- When $\mathcal{X}$ is a finite set, the FBST does not reject H_0 if and only if

$$\left(\mathbf{b}(\mathcal{X})\hat{\boldsymbol{\beta}} - \mu(\mathcal{X})\right)' \Sigma(\mathcal{X}, \mathcal{X})^{-1} \left(\mathbf{b}(\mathcal{X})\hat{\boldsymbol{\beta}} - \mu(\mathcal{X})\right) \leq q_{(1-\alpha)}(\chi^2_{|\mathcal{X}|}),$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{b}(\mathcal{X})' \Sigma(\mathcal{X}, \mathcal{X})^{-1} \mathbf{b}(\mathcal{X}))^{-1} \mathbf{b}(\mathcal{X})' \Sigma(\mathcal{X}, \mathcal{X})^{-1} \mu(\mathcal{X})$ and $|\mathcal{X}|$ is the size of $\mathcal{X}$.

- When $\mathcal{X}$ is an infinite set, the FBST does not reject H_0 if and only if

$$\bar{\mathbf{y}}_{\mathbf{X}^*}' \mathbf{M} \bar{\mathbf{y}}_{\mathbf{X}^*} \leq c_\alpha, \quad \mathbf{M} = D_{n_{\mathbf{X}^*}} - D_{n_{\mathbf{X}^*}} \mathbf{b}(\mathbf{X}^*) (\mathbf{b}(\mathbf{X}^*)' D_{n_{\mathbf{X}^*}} \mathbf{b}(\mathbf{X}^*))^{-1} \mathbf{b}(\mathbf{X}^*)' D_{n_{\mathbf{X}^*}}.$$

Before presenting the FBST for the pragmatic version of Equation (7), we specify $\mathbb{H}$ and provide the infimum when the dissimilarity function in Equation (6) is the L^2 distance in the probability space of $\mathbf{X}$. The hypothesis space $\mathbb{H}$ is such that

$$h \in \mathbb{H} \iff \mathbb{E}_{\mathbf{X}}(h^2) = \int_{\mathcal{X}} h(\mathbf{x})^2 d\mathbb{P}(\mathbf{x}) < \infty. \quad (8)$$

As for the infimum, it is described in the following Lemma:

Lemma 1 (Infimum of the dissimilarity on the linear model set). Let Equation (8) denote the hypothesis space and H_0 be the hypothesis in Equation (7). If $d(h_0, h) := \sqrt{\mathbb{E}_{\mathbf{X}}[(h_0 - h)^2]}$, then $d(H_0, h) = d(\mathbf{b} \times \tilde{\boldsymbol{\beta}}_h, h)$, $\forall h \in \mathbb{H}$, where

$$\tilde{\boldsymbol{\beta}}_h = A_b^{-1} \times \mathbf{h}_b, \quad A_b = \begin{pmatrix} \mathbb{E}[b_1^2(\mathbf{X})] & \mathbb{E}[b_2(\mathbf{X})b_1(\mathbf{X})] & \cdots & \mathbb{E}[b_k(\mathbf{X})b_1(\mathbf{X})] \\ \mathbb{E}[b_1(\mathbf{X})b_2(\mathbf{X})] & \mathbb{E}[b_2^2(\mathbf{X})] & \cdots & \mathbb{E}[b_k(\mathbf{X})b_2(\mathbf{X})] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[b_1(\mathbf{X})b_k(\mathbf{X})] & \mathbb{E}[b_2(\mathbf{X})b_k(\mathbf{X})] & \cdots & \mathbb{E}[b_k^2(\mathbf{X})] \end{pmatrix},$$

$$\mathbf{h}_b' = (\mathbb{E}[h(\mathbf{X})b_1(\mathbf{X})], \mathbb{E}[h(\mathbf{X})b_2(\mathbf{X})], \cdots, \mathbb{E}[h(\mathbf{X})b_k(\mathbf{X})]).$$

Theorem 2 (FBST of the pragmatic version of H_0). Let $\mathbb{H}$ be given by Equation (8) and define $d(h_0, h) := \sqrt{\mathbb{E}_{\mathbf{X}}[(h_0 - h)^2]}$. Assume that $\sum_{\mathbf{x} \in \mathbf{X}^*} \mathbb{P}(\mathbf{x}) > 0$. Then,

- When $\mathcal{X}$ is a finite set, the FBST does not reject $Pg(H_0)$ if and only if

$$\exists s \in (0, 1) : \quad 1 - \mu(\mathcal{X})' \left(\frac{\varepsilon^2}{1-s} \mathbf{N}^{-1} + \frac{1}{s} \Sigma(\mathcal{X}, \mathcal{X}) q_{(1-\alpha)}(\chi^2_{|\mathcal{X}|}) \right) \mu(\mathcal{X}) < 0,$$

where $\mathbf{N} := D_{\mathbb{P}(\mathcal{X})} \left[\mathbb{I} - \mathbf{b}(\mathcal{X}) (\mathbf{b}(\mathcal{X})' D_{\mathbb{P}(\mathcal{X})} \mathbf{b}(\mathcal{X}))^{-1} \mathbf{b}(\mathcal{X})' D_{\mathbb{P}(\mathcal{X})} \right]$ and $D_{\mathbb{P}(\mathcal{X})}$ is a diagonal matrix formed by the vector $\mathbb{P}(\mathbf{x}), \mathbf{x} \in \mathcal{X}$.

- When $\mathcal{X}$ is an infinite set, the FBST does not reject $Pg(H_0)$ if and only if

$$\exists s \in (0, 1) : \quad 1 - \bar{\mathbf{y}}_{\mathbf{X}^*}' \left(\frac{\varepsilon^2}{1-s} \mathbf{M}^{-1} + \frac{1}{s} D_{n_{\mathbf{X}^*}} c_\alpha \right) \bar{\mathbf{y}}_{\mathbf{X}^*} < 0,$$

where $\mathbf{M} := D_{\mathbb{P}(\mathbf{X}^*)} \left[\mathbb{I} - \mathbf{b}(\mathbf{X}^*) (\mathbf{b}(\mathbf{X}^*)' D_{\mathbb{P}(\mathbf{X}^*)} \mathbf{b}(\mathbf{X}^*))^{-1} \mathbf{b}(\mathbf{X}^*)' D_{\mathbb{P}(\mathbf{X}^*)} \right]$ and $D_{\mathbb{P}(\mathbf{X}^*)}$ is a diagonal matrix formed by the vector $\mathbb{P}(\mathbf{x}), \mathbf{x} \in \mathbf{X}^*$.

In the infinite case of $\mathcal{X}$ in Theorem 2, an appealing choice for the distribution of $\mathbf{X}$ is based on the posterior of a Dirichlet process [14]. For a concentration parameter τ and a centering distribution π , set a Dirichlet process $P \sim DP(\tau, \pi)$ such that $\mathbf{X}|P \sim P$. Then,

$$P|\mathbf{x} \sim DP\left(\frac{\tau\pi}{\tau+n} + \frac{\sum_{i=1}^n \delta_{x_i}}{\tau+n}, \tau+n\right) \implies \mathbb{P}(X_{new} \in A|\mathbf{X}=\mathbf{x}) = \frac{\tau\pi(A)}{\tau+n} + \frac{\sum_{i=1}^n \mathbb{I}(x_i \in A)}{\tau+n},$$

where $\delta_{x_i} = \mathbb{I}(x_i \in A)$ for $A \subseteq \mathcal{X}$. With this choice, one can ensure that positive probability will always be assigned to all $x \in \mathbf{X}^*$. Moreover, τ can leverage the weight of the prior on the FBST, with higher values of τ leading to a higher chance of not rejecting $Pg(H_0)$.

4. Application: Water Droplet Experiment

The dataset from Duguid [15] provides a setting where small water droplets (ranging from 3 to 9 μm) are free falling through a tube that keeps factors such as temperature and humidity constant. As a droplet falls, a camera takes a picture at every 0.5 s, ceasing activity after 7 s. One of the objectives of the study was to evaluate Fick's law, which in this setting implies that—when time is a covariate—the decrease in radius of the droplet can be described through a linear model. The two hypotheses of interest are

$$\begin{cases} H_0^1 : g(t) = \beta_0 + \beta_1 t, & \forall t \in \{0s, 0.5s, \dots, 7s\}, \quad (\beta_0, \beta_1) \in \mathbb{R}^2; \\ H_0^2 : g(t) = \beta_0, & \forall t \in \{0s, 0.5s, \dots, 7s\}, \quad \beta_0 \in \mathbb{R}; \end{cases}$$

with the first hypothesis testing the validity of Fick's law for this case and the second one verifying if time can be removed as a covariate.

We use the GP in Section 2.2 to model the data, with the following prior settings:

$$\sigma^2 = 0.01, \quad m(t) = \frac{3+9}{2} = 6, \forall t \in \mathcal{X}, \quad K(t_1, t_2) = \exp\left\{-\frac{1}{2}\|t_1 - t_2\|_2\right\}, (t_1, t_2) \in \mathcal{X}.$$

As shown in Figure 2a, this choice leads to functions that obey the 3–9 μm restriction without becoming too restrictive as a consequence. In Figure 2b, we observe that the posterior draws resemble a linear model except on $t = 0$, due to the missing observation.

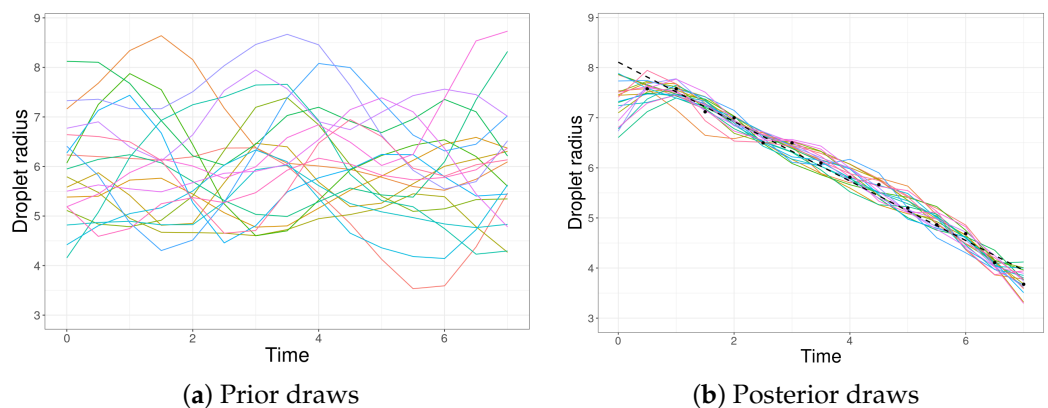


Figure 2. GP draws of the (a) prior and (b) posterior for the water droplet data. The colored curves represent each draw, the black dots are the observed data and the dashed line is the least squares estimate of the linear model.

In Table 1, we present the e-value for both hypotheses of interest, assuming that $\mathcal{X}$ is either finite or infinite. Since small e-values provide strong evidence against H_0 , with $\alpha = 0.05$ we conclude that both H_0^1 and H_0^2 should be rejected, i.e., Fick's law would fail.

Table 1. e-value of H_0 under finite and infinite $\mathcal{X}$ for the water droplet experiment.

Assumption on t	Hypothesis	
	$H_0^1 : g(t) = \beta_0 + \beta_1 t$	$H_0^2 : g(t) = \beta_0$
t is discrete and finite	0.0446	0
t is continuous	0.0068	0

While this analysis shows that Fick's law is not exactly valid, it might still provide an adequate approximation, motivating the use of pragmatic hypotheses. This requires setting the threshold ε , which is detailed below.

In the original experiment, the radius of the droplets was obtained indirectly through Stoke's law, that is,

$$V_T(t) = \frac{g(t)^2}{K_s} \implies g(t) = \sqrt{V_T(t) \times K_s}, \quad (9)$$

where V_T is the terminal velocity and $K_s = 8.446$. Since the mean velocity (V_M) was used in (9) instead of V_T , there are two sources of measurement error: the estimate of V_M (maximum error of $\delta = 0.14$) and switching V_T for V_M in (9) (maximum error of $\eta = 0.3555$ ([9], Example 1.3)). We conclude that the margin of error of the radius is

$$\begin{aligned} \epsilon &:= \max_{t \in T} |g(t) - y(t)| = \max_{t \in T} \left\{ \left| \sqrt{K_s V_T(t)} - y(t) \right| \right\} \\ &= \max_{t \in T} \left\{ \left| \sqrt{K_s (V_M(t) - \delta - \eta)} - y(t) \right|, \left| \sqrt{K_s (V_M(t) + \delta + \eta)} - y(t) \right| \right\} \approx 0.6218. \end{aligned}$$

While ϵ relates to the l_∞ distance, Lemma 1 uses the l_2 distance. To obtain an estimate of the latter from the former, we use Proposition 6.11 of Folland [16], which implies that

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y(x_i) - g(x_i))^2} \leq \sqrt{\frac{1}{n}} \epsilon \implies \max_{i \in \{1, 2, \dots, n\}} |y(x_i) - g(x_i)| \leq \epsilon,$$

thus $\varepsilon \approx 0.6218 / \sqrt{15} \approx 0.1606$.

Table 2 presents the e-values for the pragmatic hypotheses $Pg(H_0^i, d, 0.1606)$, $i \in \{1, 2\}$. We assume either that $\mathcal{X} = \{0, 0.5, 1, \dots, 7\}$ (original setting, discrete uniform) or that $t|P \sim P$, with $P \sim DP(1, U(0, 7))$ (continuous uniform as centering distribution). Contrary to Table 1, the first hypothesis is not rejected, demonstrating that Fick's law provides a good approximation of the phenomenon.

Table 2. e-value of $Pg(H_0, d, 0.1606)$ under finite and infinite $\mathcal{X}$ for the water droplet experiment.

Assumption on t	Original Hypothesis	
	$H_0^1 : g(t) = \beta_0 + \beta_1 t$	$H_0^2 : g(t) = \beta_0$
$t \in \{0, 0.5, 1, \dots, 7\}$	1	0
$t P \sim P, \quad P \sim DP(1, U(0, 7))$	1	0

5. Discussion

Regarding the results of the application (Section 4), we believe to have demonstrated the importance of using pragmatic hypotheses whenever reasonable. While choosing ε is not a simple task in nonparametric settings, there are strategies available for deriving it [9]. Furthermore, while the e-value is not a measure of evidence against H_0^c [3], combining it with a pragmatic hypothesis allows one to perform the Generalized FBST (GFBST, [17]), which can discriminate “evidence of absence” from “absence of evidence”, along with many other desirable properties.

One of the main limitations of this work is in the strategy of performing variable selection. While the aforementioned GFBST allows for multiple testing without the necessity of correcting α , variable selection is only possible through Equation (7) if the linear model hypothesis is not rejected. Therefore, one future research direction is developing tests that evaluate conditional independence without assuming a specific functional form for the relationship between variables.

Author Contributions: Conceptualization, all authors; methodology, R.F.L.L. and R.B.S.; software, R.F.L.L.; validation, R.F.L.L.; formal analysis, R.F.L.L.; investigation, J.M.S.; resources, R.F.L.L.; data curation, R.F.L.L.; writing—original draft preparation, R.F.L.L.; writing—review and editing, all authors; visualization, R.F.L.L.; supervision, R.B.S. and J.M.S.; project administration, R.B.S.; funding acquisition, Interinstitutional Graduate Program in Statistics UFSCar-USP. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001. This research was funded by FAPESP (grants 2019/11321-9 and CEPID CeMEAI 2013/07375-0) and CNPq (grants 309607/2020-5, 422705/2021-7 and PQ 303290/2021-8).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and the functions used in this study are available in GitHub at <https://github.com/rflassance/lmFBST> (accessed on 24 June 2025). These data were derived from the following resource available in the public domain: https://scholarsmine.mst.edu/cgi/viewcontent.cgi?params=/context/masters_theses/article/6294 (page 42, accessed on 15 May 2024).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

FBST	Full Bayesian Significance Test
GFBST	Generalized Full Bayesian Significance Test
GP	Gaussian Process
HPD	Highest Posterior Density
WRSS	Weighted Residual Sum of Squares

Appendix A. Proofs

Proof of Theorem 1. The proof is done in parts:

Finite $\mathcal{X}$. Since $\mathcal{X}$ is finite, Equation (2) is the HPD. Therefore, if $\exists \beta \in \mathbb{R}^k$ such that

$$(\mathbf{b}(\mathcal{X})\beta - \mu(\mathcal{X}))'\Sigma(\mathcal{X}, \mathcal{X})^{-1}(\mathbf{b}(\mathcal{X})\beta - \mu(\mathcal{X})) \leq q_{(1-\alpha)}(\chi^2_{|\mathcal{X}|}), \quad (\text{A1})$$

then the FBST does not reject H_0 . Derivating the left side of (A1) in terms of β , we observe that $\hat{\beta}$ minimizes such expression. ■

Infinite $\mathcal{X}$. In this case, the FBST does not reject the hypothesis iff $\exists \beta \in \mathbb{R}^k$ such that $WRSS(\mathbf{b} \times \beta) \leq c_\alpha$. If $\tilde{\beta} = (\mathbf{b}(X^*)'D_{n_{X^*}}\mathbf{b}(X^*))^{-1}\mathbf{b}(X^*)'D_{n_{X^*}}\bar{\mathbf{y}}_{X^*}$, this is equivalent to not rejecting H_0 iff

$$(\bar{\mathbf{y}}_{X^*} - \mathbf{b}(X^*)\tilde{\beta})'D_{n_{X^*}}(\bar{\mathbf{y}}_{X^*} - \mathbf{b}(X^*)\tilde{\beta}) = \bar{\mathbf{y}}_{X^*}'M\bar{\mathbf{y}}_{X^*} \leq c_\alpha,$$

since $\tilde{\beta}$ is the weighted least squares estimate of β . ■

□

Proof of Lemma 1. The proof is found in (Lassance et al. [9], Theorem 2). □

Proof of Theorem 2. The proof is done in parts:

Finite $\mathcal{X}$. Lemma 1 implies that

$$\tilde{\beta}_h = \left(\mathbf{b}(\mathcal{X})' D_{\mathbb{P}(\mathcal{X})} \mathbf{b}(\mathcal{X}) \right) \mathbf{b}(\mathcal{X})' D_{\mathbb{P}(\mathcal{X})} h(\mathcal{X}),$$

and thus

$$\begin{aligned} d(H_0, h(\mathcal{X})) &= \sqrt{\sum_{x \in \mathcal{X}} \mathbb{P}(x) (h(x) - \mathbf{b}(x)' \tilde{\beta}_h)^2} \\ &= \sqrt{(h(\mathcal{X}) - \mathbf{b}(\mathcal{X}) \tilde{\beta}_h)' D_{\mathbb{P}(\mathcal{X})} (h(\mathcal{X}) - \mathbf{b}(\mathcal{X}) \tilde{\beta}_h)} = \sqrt{h(\mathcal{X})' N h(\mathcal{X})} \leq \varepsilon. \end{aligned}$$

Since the HPD is given by Equation (2), the FBST does not reject $Pg(H_0)$ if and only if

$$\begin{cases} h \in \mathbb{H} : d(H_0, h(\mathcal{X})) = \sqrt{h(\mathcal{X})' N h(\mathcal{X})} \leq \varepsilon \\ h \in \mathbb{H} : (h(\mathbf{X}) - \mu(\mathcal{X}))' \Sigma(\mathcal{X}, \mathcal{X})^{-1} (h(\mathcal{X}) - \mu(\mathcal{X})) \leq q_{1-\alpha}(\chi^2_{|\mathcal{X}|}) \end{cases} \quad (\text{A2})$$

are intersecting ellipsoids. From Proposition 2 of Gilitschenski and Hanebeck [18], the ellipsoids intersect if and only if

$$\exists s \in (0, 1) : 1 - \mu(\mathcal{X})' \left(\frac{\varepsilon^2}{1-s} \mathbf{N}^{-1} + \frac{1}{s} \Sigma(\mathcal{X}, \mathcal{X}) q_{(1-\alpha)}(\chi^2_{|\mathcal{X}|}) \right) \mu(\mathcal{X}) < 0.$$

■

Infinite $\mathcal{X}$. The FBST does not reject $Pg(H_0)$ if

$$\inf_{h \in \text{HPD}} d(H_0, h)^2 = \inf_{h \in \text{HPD}} \inf_{h_0 \in H_0} d(h_0, h)^2 = \inf_{h_0 \in H_0} \inf_{h \in \text{HPD}} \int_{\mathcal{X}} (h_0(x) - h(x))^2 d\mathbb{P}_{\mathbf{X}}(x) \leq \varepsilon^2. \quad (\text{A3})$$

The ellipsoid $G := \{z \in \mathbb{R}^{n^*} : (z - \bar{\mathbf{y}}_{X^*})' D_{n_{X^*}} (z - \bar{\mathbf{y}}_{X^*}) \leq c_\alpha\}$ is such that, for any function $h(\cdot)$ where $\exists z \in G : h(\mathbf{X}^*) = z$, we can conclude that $h \in \text{HPD}$. Therefore, the HPD contains functions that are linear outside of $\mathbf{X}^*$, and thus

$$\begin{aligned} \inf_{h_0 \in H_0} \inf_{h \in \text{HPD}} \int_{\mathcal{X}} (h_0 - h(x))^2 d\mathbb{P}_{\mathbf{X}}(x) &= \inf_{h_0 \in H_0} \inf_{z \in G} \sum_{i=1}^{n^*} (z_i - h_0(\mathbf{X}_{i,\cdot}^*))^2 \mathbb{P}(\mathbf{X}_{i,\cdot}^*) \\ &= \inf_{z \in G} \inf_{h_0 \in H_0} (z - h_0(\mathbf{X}^*))' D_{\mathbb{P}(\mathbf{X}^*)} (z - h_0(\mathbf{X}^*)) \\ &= \inf_{z \in G} (z - b(\mathbf{X}^*) \hat{\beta}_z)' D_{\mathbb{P}(\mathbf{X}^*)} (z - b(\mathbf{X}^*) \hat{\beta}_z), \end{aligned}$$

where $\hat{\beta}_z = (b(\mathbf{X}^*)' D_{\mathbb{P}(\mathbf{X}^*)} b(\mathbf{X}^*))^{-1} b(\mathbf{X}^*)' D_{\mathbb{P}(\mathbf{X}^*)} z$, thus $\inf_{h \in \text{HPD}} d(H_0, h)^2 = \inf_{z \in G} z' \mathbf{M} z$. Therefore, the FBST does not reject H_0 if the ellipsoids G and $\{z \in \mathbb{R}^{n^*} : z' \mathbf{M} z \leq \varepsilon\}$ intersect, which can be verified through Proposition 2 of Gilitschenski and Hanebeck [18]. ■

□

References

1. Kershaw, J.; Kashikura, K.; Zhang, X.; Abe, S.; Kanno, I. Bayesian technique for investigating linearity in event-related BOLD fMRI. *Magn. Reson. Med.* **2001**, *45*, 1081–1094. [CrossRef] [PubMed]
2. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2002**, *64*, 583–639. [CrossRef]

3. de Bragança Pereira, C.A.; Stern, J.M.; Wechsler, S. Can a significance test be genuinely Bayesian? *Bayesian Anal.* **2008**, *3*, 79–100. [\[CrossRef\]](#)
4. Corrêa Filho, F.P.T. Nonparametric Tests for Pólya Trees: Non-Parametric Versions of the FBST. Master's Thesis, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil, 2024.
5. Liu, Z.; Li, Z.; Wang, J.; He, Y. Full Bayesian Significance Testing for Neural Networks. *arXiv* **2024**, arXiv:2401.13335. [\[CrossRef\]](#)
6. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2006.
7. Esteves, L.G.; Izbicki, R.; Stern, J.M.; Stern, R.B. Pragmatic Hypotheses in the Evolution of Science. *Entropy* **2019**, *21*, 883. [\[CrossRef\]](#)
8. Mulder, J. Bayesian Testing of Linear Versus Nonlinear Effects Using Gaussian Process Priors. *Am. Stat.* **2023**, *77*, 1–11. [\[CrossRef\]](#)
9. Lassance, R.F.; Izbicki, R.; Stern, R.B. Adding imprecision to hypotheses: A Bayesian framework for testing practical significance in nonparametric settings. *Int. J. Approx. Reason.* **2025**, *178*, 109332. [\[CrossRef\]](#)
10. Johnson, R.; Wichern, D. *Applied Multivariate Statistical Analysis*, 6th ed.; Prentice Hall India Learning Private Limited: Delhi, India, 2012; p. 163.
11. Oakley, J. Eliciting Gaussian process priors for complex computer codes. *J. R. Stat. Soc. Ser. D (Stat.)* **2002**, *51*, 81–97. [\[CrossRef\]](#)
12. Wang, X.; Berger, J.O. Estimating Shape Constrained Functions Using Gaussian Processes. *SIAM/ASA J. Uncertain. Quantif.* **2016**, *4*, 1–25. [\[CrossRef\]](#)
13. Davies, R.B. Algorithm AS 155: The Distribution of a Linear Combination of χ^2 Random Variables. *Appl. Stat.* **1980**, *29*, 323. [\[CrossRef\]](#)
14. Ferguson, T.S. A Bayesian Analysis of Some Nonparametric Problems. *Ann. Stat.* **1973**, *1*, 209–230. [\[CrossRef\]](#)
15. Duguid, H.A. A Study of the Evaporation Rates of Small Freely Falling Water Droplets. Master's Thesis, Missouri University of Science and Technology, Rolla, MO, USA, 1969.
16. Folland, G. *Real Analysis: Modern Techniques and Their Applications*, 2nd ed.; Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts; Wiley: Hoboken, NJ, USA, 1999; p. 186.
17. Esteves, L.G.; Izbicki, R.; Stern, J.M.; Stern, R.B. Logical coherence in Bayesian simultaneous three-way hypothesis tests. *Int. J. Approx. Reason.* **2023**, *152*, 297–309. [\[CrossRef\]](#)
18. Gilitschenski, I.; Hanebeck, U.D. A robust computational test for overlap of two arbitrary-dimensional ellipsoids in fault-detection of Kalman filters. In Proceedings of the 2012 15th International Conference on Information Fusion, Singapore, 9–12 July 2012; pp. 396–401.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.